In the name of Allah, the Most Gracious, the Most Merciful













Faculté de Médecine
1ière année Med et Ch/D
Biostatistique

Année 2014/2015

Série 6 : Estimation et tests statistiques

ESTIMATION

Exercice 1

Dans l'analyse du sang d'un échantillon de 100 malades pris au hasard dans une population de personnes hospitalisées pour des anomalies sanguines, on a relevé le poids de calcium X et proposé les résultats suivants : $\sum x_i = 12000\text{mg}$; $\sum x_i^2 = 1449900\ \text{mg}^2$

1) Estimer la moyenne, la variance et l'écart-type d'une mesure pour individu de la population
2) Donner l'intervalle de confiance à 95%, puis à 99% du poids moyen de calcium pour l'ensemble des malades.

Exercice 2

On a pris au hasard 16 sujets parmi la population d'hypertendus. Chez chacun d'eux, on a mesuré l'aldolase (ALD) en U.I par litre.

Les résultats sont :

2,5 3,2 0,8 3,2 3,2 3,2 2,5 4,6 3,6 6,1 2,9 1,0 5,1 4,7 4,7 3,2

1) Donner les estimations de la moyenne et de la variance de l'ALD dans la population des Hypertendus
2) Calculer les intervalles de confiance pour la moyenne et l'écart type de l'ALD chez les hypertendus.

Exercice 3

On a relevé les âges (en années) de 857 consultants pris au hasard dans un service hospitalier, les résultats ont permis de calculer $\bar{x} = 27{,}445$ *et* $S_x^2 = 31{,}303$.

Calculer les intervalles de confiance à 90% de la moyenne et de la variance de l'âge des consultants de ce service.

Exercice 4

On a mesuré chez 50 adultes le taux d'acide urique et on a obtenu une moyenne de 47,3 mg/l et un écart-type de 1,85 mg/l.

1) Donner les intervalles de confiance aux niveaux 95% et 90% pour le taux moyen d'acide urique dans la population. Comparer ces deux intervalles.
2) On suppose que la variance dans la population est égale à 2,5 m/g^2. Donner le nombre d'adultes n_0 que l'on doit examiner pour que l'intervalle de confiance soit [46,9 ; 48,1] au risque de 5%, puis au risque de 1%. Conclure.

Exercice 5

On étudie le taux d'un gaz nocif dans l'atmosphère pour un volume donné.

1) Pour cela on extrait un échantillon de taille 10 sur lequel on observe une moyenne de 50 et une variance de 100.
   a) Quel est l'intervalle de confiance à 95%, pour le taux moyen de gaz dans l'atmosphère ? Quelle hypothèse doit-on émettre ?
   b) On suppose que la variance du taux de gaz dans l'atmosphère est de 100. Quel serait alors cet l'intervalle de confiance à 95% ?
2) On dispose d'un échantillon de taille 100 de moyenne de 48 et de variance de 90. Quel est alors l'intervalle de confiance pour le taux moyen de gaz dans l'atmosphère au risque de 5%?

Exercice 6

On a mesuré la pression sanguine chez 32 chèvres. L'intervalle de confiance à 95% de la pression moyenne est [4,7 ; 5,5] en cm de mercure (Hg).

1. Déterminer la pression moyenne correspondant aux 32 chèvres.
2. Quel est le nombre de chèvres qu'il aurait fallu utiliser pour obtenir une précision d'au moins 2% ? On supposera que la moyenne et la variance restent inchangées.

Exercice 7

Une étude sur un échantillon de 100 souris d'une certaine race a montré que la présence de cancers spontanés est de 25%

1. Donner l'intervalle de confiance au risque de 5% pour la proportion de cancers spontanés dans la population.
2. Même question si 25% est obtenu à partir de 100000 souris
3. Quelle taille d'échantillon doit-on prendre pour avoir une précision d'au moins 1% ?

Exercice 8

Un caractère a été estimé comme touchant entre 19,3% et 20,1% d'une population avec un risque de se tromper de 1%. Quel a été le nombre de sujets nécessaires à ce sondage ?

Exercice 9

Un service de santé local veut estimer la prévalence de la tuberculose chez les moins de 5 ans dans la zone de son ressort.

Combien faut-il inclure d'enfants dans l'échantillon pour que la prévalence soit connue avec une erreur maximale de 2% par rapport à la valeur exacte, pour un niveau de confiance de 95% et en admettant que la proportion réelle est aux alentours de 20% ?

TESTS DE CONFORMITE

Exercice 1

Le taux moyen de glycémie pour une certaine population est de 0,9 g/l.

1. Sur un échantillon de 37 sujets, le taux moyen de glycémie est de 1.2 g/l avec un écart-type de 0.4 g/l. Peut-on considérer, au seuil de confiance de 0.95, que l'échantillon est représentatif de la population ?
2. Quelle serait votre réponse si on ne dispose que de 26 sujets ? (On supposera que le taux de glycémie suit une loi normale).

Exercice 2

Dans un échantillon de 16 enfants on a dénombré 11 vaccinés par le B.C.G. Sachant que le pourcentage de vaccinés dans le pays est de 75%, peut-on affirmer que cet échantillon est représentatif de la population ?

Exercice 3

Afin de tester une solution toxique, on l'injecte à un groupe de 80 souris. On admet que l'injection est mortelle dans 80% des cas. Le fait que 20 souris ne soient pas mortes, est-il compatible, au niveau de confiance de 99%, avec l'hypothèse faite sur le taux de mortalité ? Procéder par 3 méthodes

Exercice 4

On a croisé 2 races de plantes différant par 2 caractères A et B. La première génération est homogène, la seconde fait apparaître 4 types de plantes dont le phénotype est noté AB, Ab, aB, ab.

Si les caractères se transmettent selon les lois de Mendel, les proportions théoriques des 4 phénotypes sont respectivement : 9/16, 3/16, 3/16, 1/16.

Dans une expérience, un échantillon de 160 plantes a donné : 100 pour AB, 18 pour Ab, 24 pour aB et 18 pour ab.

Cette répartition est-elle conforme aux lois de Mendel ?

Exercice 5

On sait que pour l'ensemble de l'Europe, les pourcentages des 4 groupes sanguins s'établissent ainsi : 45% pour O, 35% pour A, 16% pour B et 4% pour AB.

Un échantillon de 100 individus est prélevé au hasard, dans une zone montagneuse, 35 sont du groupe O, 35 de A, 20 de B et 10 de AB. Peut-on penser qu'il y a conformité entre ces résultats et ceux établis pour l'ensemble de l'Europe au risque de 5% ?

TESTS D'HOMOGENEITE

Exercice 1

On prélève au hasard 2 échantillons d'individus, l'un en milieu urbain et l'autre en mili rural. On évalue le rythme cardiaque au repos, X, pour chaque individu. On obtient :

	Milieu urbain	Milieu rural
Taille	100	140
Moyenne	80	77
Variance	150	210

1. Calculer les intervalles de confiance des moyennes au taux de sécurité (T.S) de 95%.
2. Peut-on conclure que le rythme cardiaque au repos augmente significativement en mili urbain ? Faire le test approprié si nécessaire.
3. Peut-on répondre directement à la question précédente au risque de 1% ?
4. En supposant que les 2 échantillons ont la même taille,n ; et que les moyennes et variances restent inchangées, trouver la valeur minimale de n qui permettrait d'affirn que les rythme cardiaques sont, en moyenne, significativement différents. Répondre utilisant a) le test de l'écart réduit
   b) le recouvrement des intervalles de confiance
5. Comparer les 2 tailles obtenues. Quelle méthode est-il plus raisonnable de choisir ?

Exercice 2

Les statistiques de la maternité d'un hôpital donnent les poids des nouveau-nés par sexe en

	Garçons	Filles
Taille	15	16
Moyenne	3400	3360
Ecart type	380	360

Peut-on déduire de ces données une différence significative de poids suivant le sexe des nouveau-nés au taux de sécurité de 90% ?
Quelle hypothèse faut-il émettre sur la population ?

Exercice 3

Une étude sur l'effet de la streptomycine sur la tuberculose pulmonaire a porté sur malades dont 55 traités et 52 témoins (non traités). On a observé 4 et 14 décès respectivem

1. Calculer l'intervalle de confiance du pourcentage de décès chez les traités et les tém au risque de 5%.
2. En comparant ces intervalles, peut-on considérer que la streptomycine a un effet sur c maladie ?

Exercice 4

L'étude expérimentale d'un médicament a été pratiqué sur 100 malades divisés, par tirage sort en 2 groupes A et B.

- Le groupe A, composé de 60 malades, a absorbé le médicament
- Le groupe B composé de 40 malades, n'a absorbé qu'un placebo (produit inactif extérieurement identique au médicament)

Les résultats sont les suivants : Groupe A : 40 malades guéris
Groupe B : 20 malades guéris

Peut-on conclure à l'efficacité du médicament au risque de 5% ?

TESTS D'AJUSTEMENT

Exercice 1 :

Au cours d'un jeu de " pile ou face ", sur 100jets consécutifs de 4 pièces identiques, on obtient les résultats suivants :

Nombre de piles	0	1	2	3	4
Nombre de jets	12	25	36	25	2

1. a/ Quelle est la moyenne du nombre de piles par jet de 4 pièces ?
   b/ Soit X le nombre de piles par jet de 4 pièces. En supposant que X suit une loi binomiale, déduire la probabilité élémentaire p d'obtenir pile pour un jet d'une pièce.
   c/ Comparer les résultats expérimentaux avec les résultats théoriques déduits de cette loi binomiale. Conclusion ?
2. Comparer la distribution observée à celle déduite d'une loi binomiale où p=1/2.

Exercice 2

A partir de la vente de 100 postes de télévision ayant fonctionné le même nombre d'heures, pendant une année, on a pu établir le tableau suivant contenant le nombre de réparation de ces téléviseurs :

Nombre de réparations	0	1	2	3	4
Nombre de téléviseurs	61	30	7	2	0

1. a/ Calculer la moyenne observée $\bar{x}$.
   b/ Peut-on ajuster la distribution observée à une loi de Poisson de paramètre $\bar{x}$?
2. Peut-on l'ajuster à la loi de Poisson de paramètre $\lambda = 1$?

Exercice 3 :

On a mesuré la taille en cm de 200 étudiants, et on a obtenu les résultats suivants :

taille	[134,142[	[142,150[	[150,158[	[158,166[	[166,174[	[174,182[
effectif	16	44	60	44	26	10

1. Déterminer la moyenne m et l'écart-type σ.
2. Peut-on considérer que l'échantillon étudié est issu d'une population dont la moyenne est de 160cm, au niveau de confiance de 95% ?
3. Peut-on ajuster cette répartition observée à la loi Normale $N(m, \sigma^2)$, au risque de 10% ?

TESTS D'INDEPENDANCE - TESTS DE CORRELATION

Exercice 1 :
On se demande si la localisation précise d'un cancer de l'estomac est liée ou non au sexe du malade. Pour cela, on dispose des renseignements suivants, portant sur 298 cas :

Localisation / Sexe	Région du pylore	Corps de L'estomac	Région du cardia
Hommes	53	66	75
Femmes	48	33	23

La localisation du cancer de l'estomac, et le sexe sont-ils liés au risque de 5% ?

Exercice 2:
On veut comparer les réactions produites par deux vaccins B.C.G désignés par A et B.
Un groupe de 348 enfants a été divisé par tirage au sort en deux séries qui ont été vaccinés l'une par A, l'autre par B. La réaction a été lue ensuite par une personne ignorant le vaccin utilisé. Les résultats sont :

Réaction / Vaccin	Légère	Moyenne	Ulcération	Abcès
A	12	156	8	1
B	29	135	6	1

Existe-t-il une relation entre les réactions et le type de vaccin utilisé au risque de 5% ?

Exercice 3 :
Pour un échantillon de 400 femmes, on dispose de courbes donnant la température journalière pendant un cycle mensuel où il y a eu conception.
Ces courbes de température comportent une phase de température basse et une phase de température élevée, séparées par une montée thermique. La montée thermique peut être classée " lente " ou " rapide ".
Les 400 conceptions se sont terminées, soit par une naissance, soit par un avortement spontané en cours de grossesse.
La répartition des 400 femmes en fonction de la nature de la montée thermique et de l'issue de la grossesse est donnée par le tableau suivant :

Montée thermique / Issue de la grossesse	Rapide	Lente
Avortement spontané	2	38
Naissance	78	282

On veut savoir s'il existe une liaison entre la nature de la montée thermique et l'issue de la grossesse.

1. Pour établir cette liaison, quelles proportions observées doit-on comparer ? (Préciser les valeurs numériques de ces proportions et la méthode permettant de les comparer).
2. Montrer par 2 méthodes, que la liaison entre les 2 caractères est significative au risque de 5%.

Exercice 4 :

On a observé 170 nouveau-nés dans une maternité et noté ceux qui étaient porteurs d'un angiome (ou " tache de vin ").On veut tester l'existence d'une liaison entre le caractère de la grossesse (normal ou pathologique) et la présence d'un angiome, à partir du tableau suivant :

Grossesse \ Angiome	Présence	Absence
Normale	5	135
Pathologique	2	28

Quelle est votre réponse à 95% ? et à 99% ?

Exercice 5 :

On a pris 2 mensurations X et Y sur une coquille de mollusque (X = la longueur de la coquille et Y = la largeur de la coquille) .

On a calculé pour un échantillon de 100 mollusques les quantités suivantes :

Moyenne des x = 20 ; Moyenne des y = 12

Moyenne des x^2 = 436 ; Moyenne des y^2 = 153

Moyenne des produits xy = 254,4.

1. Calculer le coefficient de corrélation linéaire entre X et Y.
2. Etablir les droites de régression de y en x et de x en y.

Exercice 6 :

On désire savoir s'il existe une liaison entre le poids de naissance d'un enfant et l'âge de la mère à l'accouchement. On a calculé les quantités suivantes :

Age de la mère (X) : Moyenne = 25 ans ; Variance = $25(\text{ans})^2$

Poids du bébé (Y) Moyenne = 3100g ; Variance = $10000\ (g)^2$

La covariance observée est de 450g.ans

1. Calculer le coefficient de corrélation.
2. Le poids du bébé et l'âge de la mère sont-ils significativement liés au risque de 5% ?
3. le poids d'un bébé dont la mère a 20 ans est-il supérieur à au poids d'un bébé dont la mère a 30 ans ?

Exercice 7:

Soit la série de 12 mesures suivantes

x_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	1,5	2,5	2	2	3	4	4	3,5	4,5	4	5	5,5

1. Calculer le coefficient de corrélation entre les variables X et Y.
2. Peut-on considérer que la corrélation est significative entre ces 2 variables au risque 1% ?
3. Trouver la droite de régression de y en x.
4. Peut-on prédire la valeur de Y pour X = 13 ?

Exercice 8 :

Pour étudier la corrélation entre 2 caractères quantitatifs A et B, on a pris 2 échantillons dans 2 populations distinctes :

Echantillon 1 : taille n_1= 400 individus , coefficient de corrélation r_1 = 0,20

Echantillon 2 : taille n_2= 625 individus , coefficient de corrélation r_2 = 0,38.

1. Peut-on affirmer qu'il y a une corrélation significative dans l'échantillon 1 ?
2. Même question pour l'échantillon 2.
3. Pensez-vous que les 2 populations dont sont tirés les deux échantillons sont identiques?

Série 5:

Rappel:

Estimation ponctuelle: pour un éch. on note la moyenne par $\bar{X}$, et la variance par S^2 : $S^2 = \frac{1}{n}\sum x_i^2 - \bar{X}^2$

Estimateur de la moyenne de la pop : $\hat{m} = \bar{X}$.

Estimateur de la variance de la pop. : $\hat{\sigma}^2 = \frac{n}{n-1} S^2$.

Estimation par intervalle de confiance

Moyenne :

σ^2 \ n	σ^2 connue	σ^2 inconnue
$n>30$, X n'importe quelle	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}$	$\bar{X} \pm \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}}$ (Student)
$n<30$, $X \rightsquigarrow N(m,\sigma^2)$	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}$ (Student)	$\bar{X} \pm \frac{S}{\sqrt{n-1}} t_{n-1}(\alpha)$

Variance :

Moyenne connue:

$$IC(\sigma^2) = \left[\frac{\sum(x_i-m)^2}{\chi^2_n(1-\frac{\alpha}{2})} ; \frac{\sum(x_i-m)^2}{\chi^2_n(\frac{\alpha}{2})}\right]$$

Moyenne inconnue:

$$IC(\sigma^2) = \left[\frac{nS^2}{\chi^2_{n-1}(1-\frac{\alpha}{2})} ; \frac{nS^2}{\chi^2_{n-1}(\frac{\alpha}{2})}\right]$$

Exercices:

Exo 1:

$n = 100$ X : v.a. : le poids de Ca

$\sum x_i = 12000\,mg$ $\sum x_i^2 = 1449900\,mg^2$

1/ $\hat{m} = \bar{X} = \frac{1}{n}\sum x_i = \frac{1}{100} 12000$

$\hat{m} = 120\,mg$

$\hat{\sigma}^2 = \frac{n}{n-1} S^2$

$S^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$

$S^2 = \frac{1}{100} \cdot 1449900 - (120)^2 = 99$

$\hat{\sigma}^2 = \frac{100}{99} \times 99$.

$\hat{\sigma}^2 = 100$.

$\hat{\sigma} = \sqrt{100} = 10$.

2/ $IC(m)$? $n = 100 > 30$ σ^2 inconnue

$$IC(m) = \left[\bar{X} - \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}} ; \bar{X} + \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}}\right]$$

$\alpha = 5\%$

$$IC_\alpha(m) = \left[120 - \frac{\sqrt{99}}{\sqrt{99}} q_{0,975} ; 120 + \frac{\sqrt{99}}{\sqrt{99}} q_{0,975}\right]$$

$IC_\alpha(m) = [118,04 ; 121,96]$

$IC_{\alpha=5\%}(m) = [118,04 ; 121,96]$.

$IC(m)$? $n = 100 > 30$ σ^2 inconnu.

$\alpha = 1\% = 0,01$.

$$IC_\alpha(m) = \left[120 - \frac{\sqrt{99}}{\sqrt{99}} q_{0,995} ; 120 + \frac{\sqrt{99}}{\sqrt{99}} q_{0,995}\right]$$

$IC_{\alpha=1\%}(m) = [117,42 ; 122,58]$.

Exo 2:

1/ $n = 16$

$\hat{m} = \bar{X} = 3,4$.

$\hat{\sigma}^2 = \frac{n}{n-1} S^2$ $S^2 = \frac{1}{n}\sum x_i^2 - \bar{X}^2$

$S^2 = 1,97$.

$\hat{\sigma}^2 = 2,10$.

$\hat{\sigma} = 1,45$.

2/ $IC(m)$? $n = 16 < 30$ σ^2 inconnue

On pose l'hypothèse $X \rightsquigarrow N(m, \sigma^2)$.

$$IC(m) = \left[\bar{X} - \frac{S}{\sqrt{n-1}} t_{n-1}(\alpha) ; \bar{X} + \frac{S}{\sqrt{n-1}} t_{n-1}(\alpha)\right]$$

$t_{15}(0{,}05) = 2{,}13 \rightarrow$ Table de Student

Quand α n'est pas donné, on le prend à 5%.

$IC_\alpha(m) = [2{,}62 ; 4{,}17]$

$IC(\sigma^2)$? m inconnue $\alpha = 5\%$

$$IC_\alpha(\sigma^2) = \left[\frac{nS^2}{\chi^2_{n-1}(1-\frac{\alpha}{2})} ; \frac{nS^2}{\chi^2_{n-1}(\frac{\alpha}{2})}\right]$$

$IC_\alpha(\sigma^2) = [1{,}15 ; 5{,}04]$ ($\sqrt{}$; $\sqrt{}$)

$IC_\alpha(\sigma) = [1{,}07 ; 2{,}24]$

Exo 3 :

$n = 857 > 30$ $\bar{X} = 27{,}445$ $S_x^2 = 31{,}303$

$\alpha = 0{,}1$.

$IC(m)$? $n > 30$. σ^2 inconnue

$$IC_\alpha(m) = \left[\bar{X} - \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}} ; \bar{X} + \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}}\right]$$

$IC_\alpha(m) = [27{,}14 ; 27{,}76]$

$IC_\alpha(\sigma^2)$? m inconnue $\alpha = 10\%$.

$$IC_\alpha(\sigma^2) = \left[\frac{nS^2}{\chi^2_{n-1}(1-\frac{\alpha}{2})} ; \frac{nS^2}{\chi^2_{n-1}(\frac{\alpha}{2})}\right]$$

Trouver les 2 quantiles $\Rightarrow$ approximer par la loi normale $n = 857$ très grand.

$$P\left[\frac{nS^2}{\chi^2_{n-1}(1-\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{nS^2}{\chi^2_{n-1}(\frac{\alpha}{2})}\right] = 1 - \alpha.$$

$$= \frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$

On va approximer par la loi normale de paramètres $m = ddl = n-1$

$= 857 - 1$

$= 856$.

et d'une variance $\sigma^2 = 2\,ddl = 2 \times 856$

$\sigma^2 = 1712$

$P(Y < y_1) = \alpha/2$

$P(Y < y_2) = 1 - \frac{\alpha}{2}$

$$P\left(\frac{Y-m}{\sigma} < \frac{y_1 - m}{\sigma}\right) = \frac{\alpha}{2}.$$

$$P\left(Y' < \frac{y_1 - 856}{\sqrt{1712}}\right) = 0{,}95.$$

$$\Phi\left(\frac{y_1 - 856}{\sqrt{1712}}\right) = 0{,}05.$$

$$\frac{y_1 - 856}{\sqrt{1712}} = -1{,}65 \Rightarrow y_1 = 787{,}72.$$

- $P(Y < y_2) = 1 - \frac{\alpha}{2}$.

$$P\left(Y < \frac{y_2 - 856}{\sqrt{1712}}\right) = 0{,}95.$$

$$\Phi\left(\frac{y_2 - 856}{\sqrt{1712}}\right) = 0{,}95.$$

$$\frac{y_2 - 856}{\sqrt{1712}} = 1{,}65.$$

$\Rightarrow y_2 = 924{,}27$.

Donc :

$$IC_\alpha(\sigma^2) = \left[\frac{857 \times 31{,}30}{924{,}27} ; \frac{857 \times 31{,}30}{787{,}72}\right]$$

$IC_\alpha(\sigma^2) = [29{,}02 ; 34{,}05]$.

Exo 4 :

1/ $n = 50$ $\bar{x} = 47{,}3$ $\sigma = 1{,}85$.

$IC(m) = ?$ avec $\alpha = 5\%$

σ^2 inconnue.

$$IC(m) = \left[\bar{X} - \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}} ; \bar{X} + \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}}\right]$$

$IC_{\alpha=5\%}(m) = [46{,}78 ; 47{,}81]$

avec $\alpha = 10\%$

$IC_{\alpha=10\%}(m) = [46{,}86 ; 47{,}73]$.

Lorsque le risque augmente, IC se réduit

2/ $\sigma^2 = 2,5$ connue.

$$IC(m) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} ; \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$$

$$IC(m) = [46,9 ; 48,1].$$

avec $\alpha = 5\%$: $q_{1-\frac{\alpha}{2}} = 1,96$.

$$\begin{cases} \bar{X} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} = 46,9 \dots (1) \\ \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} = 48,1 \dots (2) \end{cases}$$

$\ominus$

$n = 27$

avec $\alpha = 1\%$:

$n = 47$.

Lorsque le risque diminue, la taille de l'éch. ~~diminue~~ augmente.

Exo 5:

1/ $n = 100 > 30$ $\bar{X} = \hat{m} = 50$ $\hat{\sigma}^2 = 100 = s^2$

σ^2 inconnue $\alpha = 5\%$

a/ $IC_{\alpha=5\%}(m) = \left[\bar{X} - \frac{S}{\sqrt{n-1}} t_{n-1}(\alpha) ; \bar{X} + \frac{S}{\sqrt{n-1}} t_{n-1}(\alpha)\right]$

$t_{n-1}(\alpha) = 2,26$.

L'hypothèse qu'on doit émettre est que la pop. suit une loi normale.

$IC_{\alpha=5\%}(m) = [42,46 ; 57,54]$.

b/ $IC_{\alpha=5\%}(m) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} ; \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$

$q_{1-\frac{\alpha}{2}} = 1,96$.

$IC_{\alpha=5\%}(m) = [43,80 ; 56,19]$.

2/ $n = 100 > 30$. σ^2 inconnue.

$\bar{X} = 48$. $\alpha = 5\%$. $S^2 = 90$

$$IC_{\alpha=5\%}(m) = \left[\bar{X} - \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}} ; \bar{X} + \frac{S}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}}\right]$$

$$IC_{\alpha=5\%}(m) = [46,13 ; 49,87].$$

Exo 6:

1/ $n = 32 \geq 30$

$IC_{95\%}(m) = [\overset{H_1}{4,7} ; \overset{H_2}{5,5}]$

$\bar{X} = ?$

$H_1 = \bar{X} - h = 4,7 \rightarrow (1)$.

$H_2 = \bar{X} + h = 5,5 \rightarrow (2)$.

$(1) + (2) \Leftrightarrow 2\bar{X} = 10,2 \Leftrightarrow \bar{X} = 5,1$

2/ $n_0 = ?$

Pour avoir une précision d'au moins 2% $\Leftrightarrow h$

$n = 32 \geq 30$; σ inconnue

$$h = q_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n_0 - 1}}$$

$$h \leq 2\% \Leftrightarrow q_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n_0 - 1}} \leq 0,02$$

$$\Leftrightarrow n_0 \geq \left(\frac{q_{1-\frac{\alpha}{2}}}{0,02}\right)^2 S^2 + 1.$$

$$h = q_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}} \Leftrightarrow S = \frac{h\sqrt{n-1}}{q_{1-\frac{\alpha}{2}}}$$

$H_1 = \bar{X} - h \rightarrow (1)$

$H_2 = \bar{X} + h \rightarrow (2)$.

$(2) - (1) \Leftrightarrow 2h = 5,5 - 4,7$

$= 0,8$.

$\Rightarrow h = 0,4$

$n_0 \geq \left(\frac{1,96}{0,02}\right)^2 (1,13)^2 + 1 = 12264,35$

Exo 7 :

1/ $n = 100 \geq 30$.

$p = 25\%$.

$$IC_{95\%}(p_0) = \left[p \pm q_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right]$$

$$= \left[0,25 \pm 1,96 \sqrt{\frac{(0,25)(0,75)}{100}}\right]$$

$$= [0,25 \pm 0,08] = [0,16 ; 0,33].$$

2/ $n = 100\,000$

$$IC_{95\%}(p_0) = \left[0,25 \pm 1,96 \sqrt{\frac{(0,25)(0,75)}{100\,000}}\right]$$

$$= [0,25 \pm 0,00027]$$

$$= [0,2497 ; 0,2527].$$

3/ $n_0 = ?$ Pour avoir une précision d'au moins 1 % $\Leftrightarrow h \leq 1\%$

$$q_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n_0}} \leq 0,01 \Leftrightarrow n_0 \geq \left(\frac{q_{1-\frac{\alpha}{2}}}{0,01}\right)^2 (p(1-p))$$

$$\Leftrightarrow n_0 \geq \left(\frac{1,96}{0,01}\right)^2 (0,25 . 0,75)$$

$$n_0 \geq 7203.$$

Exo 8 :

$$IC_{99\%}(p_0) = [0,193 ; 0,201].$$

1/ $n = ?$

$H_1 = p - h = 0,193 \rightarrow (1).$

$H_2 = p + h = 0,201 \rightarrow (2).$

$(1) + (2) \Leftrightarrow 2p = 0,394$

$\Rightarrow p = 0,197.$

$$h = q_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

$(2) - (1) \Leftrightarrow 2h = 0,008$

$\Rightarrow h = 0,004.$

Donc : $n = \left(\frac{q_{1-\frac{\alpha}{2}}}{h}\right)^2 [p(1-p)].$

$$n = \left(\frac{2,58}{0,004}\right)^2 [(0,197)(0,803)]$$

$n = 65811$

Exo 9 :

9

10

Tests de conformité

Rappel.

Test de χ^2 :

$$\chi^2_{cal} = \sum \frac{(o_i - c_i)^2}{c_i}$$

- o_i : effectif observé
- c_i : effectif théorique
- $c_i = n p_i \quad c_i \geq 5$

Comparer χ^2_{cal} à $\chi^2_{tabulé}$

si $\chi^2_{cal} < \chi^2_{k-1}(\alpha) \Rightarrow$ On accepte H_0

k : nbre de classes

si $\chi^2_{cal} > \chi^2_{k-1}(\alpha) \Rightarrow$ On rejete H_0

Exercices:

Exo1:

1/ H_0 : l'éch. est représentatif de la pop.

$n = 37$, $\bar{x} = 1,2$, $s = 0,4$, $\alpha = 5\%$, $m = 0,9$

Intervalle de confiance de pari :

$$IP = \left[m \pm \frac{s}{\sqrt{n-1}} q_{1-\frac{\alpha}{2}}\right]$$

$$IP = \left[0,9 \pm \frac{0,4}{\sqrt{36}} 1,96\right]$$

$$IP = [0,77 \ ; \ 1,03]$$

$\bar{x} = 1,2 \notin IP \Rightarrow$ On rejette H_0

Test de l'écart réduit :

σ^2 inconnu :

$$\varepsilon = \frac{|\bar{x} - m|}{s/\sqrt{n-1}} = \frac{|1,2 - 0,9|}{0,4/\sqrt{36}} = 4,5$$

$n = 37 \rightarrow$ on compare ε au $q_{1-\frac{\alpha}{2}} = 1,96$

$\varepsilon > 1,96 \Rightarrow$ on rejette H_0

Donc l'éch. n'est pas représentatif de la pop.

2/ $IP = \left[m \pm \frac{s}{\sqrt{n-1}} t_{n-1}(\alpha)\right]$ $\quad t_{25}(0,05) = 2,06$

$$IP = \left[0,9 \pm \frac{0,4}{\sqrt{25}} 2,06\right]$$

$$IP = [0,74 \ ; \ 1,06]$$

$\bar{x} = 1,2 \notin IP \Rightarrow$ on rejette H_0

Test de l'écart réduit :

σ^2 inconnu :

$$\varepsilon = \frac{|\bar{x} - m|}{s/\sqrt{n-1}} = \frac{|1,2 - 0,9|}{0,4/\sqrt{25}} = 3,75$$

$n = 26 < 30 \rightarrow$ on compare ε au

$t_{n-1}(\alpha) = t_{25}(0,05) = 2,06$.

$\varepsilon > 2,06 \Rightarrow$ on rejette H_0

Donc l'éch. n'est pas représentatif de la pop.

Exo2:

H_0 : l'éch. est représentatif de la pop.

$n = 16 \qquad p_0 = 0,75$.

$\Rightarrow$ Test de χ^2.

On ne peut pas utiliser le test de l'écart réduit.

$n p_0 = 12 > 5$.

$n q_0 = 4 < 5$.

o_i : effectif observé : 11 vaccinés
5 non vaccinés

c_i : effectif théorique.

$c_i = n q_0 \Rightarrow C_1 = 16 \times 0,75 = 12$

$C_2 = 16 \times 0,25 = 4 < 5$

$C_2 < 5$ on calcule le χ^2 en utilisant la correction de Yates.

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(|o_i - c_i| - 1/2)^2}{c_i}$$

$\chi^2_{cal} = 0{,}082$

$\chi^2_{2-1}(0{,}05) = \chi^2_1(0{,}05) = 3{,}84$.

$\chi^2_{cal} < \chi^2_t \Rightarrow H_0$ est acceptée au risque α donc l'éch. est représentatif.

Exo 3:

H_0 : l'éch. est représentatif de la pop.

$n = 80$ $p_0 = 0{,}8$ $\alpha = 1\%$

$n = 80$ souris → 20 non mortes ; → 60 mortes Mortalité

$p = \frac{60}{80} = 0{,}75$.

Méthode 1: Intervalle de pari:

$IP = \left[p_0 \pm q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0 q_0}{n}}\right]$

$IP = \left[0{,}8 \pm 2{,}58 \sqrt{\frac{0{,}8 \times 0{,}2}{80}}\right]$

$IP = [0{,}68 ; 0{,}92] \Rightarrow P \in IP$ on accepte H_0

Méthode 2: Écart réduit:

$np_0 = 64 > 5$

$nq_0 = 16 > 5$

$\varepsilon = \frac{|p - p_0|}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{|0{,}75 - 0{,}8|}{\sqrt{\frac{0{,}8 \cdot 0{,}2}{80}}} = 1{,}12$

$q_{1-\frac{\alpha}{2}} = 2{,}58$.

$\varepsilon < q_{1-\frac{\alpha}{2}} \Rightarrow$ on accepte H_0 au risque α

Méthode 3: Test de χ^2:

$C_1 = np_0 = 64$.

$C_2 = nq_0 = 16$.

$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(O_i - C_i)^2}{C_i}$

$\chi^2_{cal} = 1{,}25$

$\chi^2_{tab} = \chi^2_{2-1}(0{,}01) = 6{,}63$.

$\chi^2_{cal} < \chi^2_{tab} \Rightarrow H_0$ est acceptée.

L'éch. est conforme à la pop.

Exo 4:

Répartition ⇒ Test de χ^2.

H_0 : La répartition est conforme aux lois de Mendel.

$C_i = np_0$ $C_1 = 160 \times 9/16 = 90$.

$C_2 = 160 \times 3/16 = 30$.

$C_3 = 160 \times 3/16 = 30$.

$C_4 = 160 \times 1/16 = 10$

	AB	Ab	aB	ab	Σ
O_i	100	18	24	18	160
C_i	90	30	30	10	160
$\frac{(O_i-C_i)^2}{C_i}$	1,11	4,8	1,2	6,4	13,51

$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(O_i - C_i)^2}{C_i} = 13{,}51$

$\chi^2_{tab} = \chi^2_{4-1}(0{,}05) = 7{,}81$

$\chi^2_{cal} > \chi^2_{tab} \Rightarrow H_0$ est rejettée au risque α.

Cette répartition n'est pas conforme aux lois de Mendel.

Rappel :

Test d'ajustement : c'est un test de comparaison de deux répartitions (théorique et observée) mais avec une répartition théorique usuelle.

Test du χ^2 :

$$\chi^2 = \sum \frac{(\theta_i - c_i)^2}{c_i} \qquad c_i = n p_i$$

(p_i : Probabilités de la loi usuelle)

$(c_i > 5)$.

- Binomiale : $P(X=i) = C_n^i \, p^i (1-p)^{n-i}$
- Poisson : $P(X=i) = \frac{e^{-\lambda}}{i!} \lambda^i$.
- Normale.

$\chi^2_{cal} < \chi^2_{tab} \Rightarrow H_0$ acceptée.

$\chi^2_{cal} > \chi^2_{tab} \Rightarrow H_0$ rejetée.

Exercices :

Exo 1 :

1/ a/ $\bar{x} = \frac{1}{N} \sum n_i x_i$

$\bar{x} = 1,8$

b/ $X \sim B(n,p)$ avec : $n = 4$.

p inconnu.

$p = \hat{p} = \frac{\bar{x}}{n}$

$\hat{p} = 0,45$

Donc : $X \sim B(4, 0,45)$.

c/ H_0 : la répartition de l'éch. peut être ajustée à la loi binomiale $B(4, 0,45)$.

• Calcul des $P_i = C_4^i (0,45)(0,55)^{4-i}$ $i = \overline{0,4}$:

$P(X=0) = C_4^0 (0,45)(0,55)^4$

$= 0,09$.

$P(X=1) = 0,3$

$P(X=2) = 0,37$

$P(X=3) = 0,2$

$P(X=4) = 0,04$

	0	1	2	3	4	Σ
P_i	0,09 / 0,0625	0,3 / 0,25	0,37 / 0,375	0,2 / 0,25	0,04 / 0,0625	1
θ_i	12	25	36	25	2	100
C_i	9 / 6,25	30 / 25	37 / 37,5	20 / 25	4 / 6,25	100
θ_i'	12	25	36	27 (3–4)		100
C_i'	9 / 6,25	30 / 25	37 / 37,5	24 (3–4) / 25	6,25	100
$\frac{(\theta_i - C_i)^2}{C_i}$	1 / 5,29	0,83 / 0	0,03 / 0,06	0,38 (3–4) / 0	2,89	2,24 / 8,24

$\chi^2_{cal} = \sum \frac{(\theta_i - c_i)^2}{c_i} = 2,24$.

$\chi^2_{thé} = \chi^2_{k-r-1}(\alpha)$ avec : r : nbre de paramètres estimé

$= \chi^2_2(0,05)$. k : nbre de classes après regroupement

$\chi^2_{thé} = 5,99$ $\alpha = 5\%$

$\chi^2_{cal} < \chi^2_{théo}$

$\Rightarrow H_0$ acceptée au risque α.

2/ H_0 : la répartition de l'éch. peut-être ajustée à la loi binomiale $B(4, 0,5)$.

$\chi^2_{cal} = 8,24$.

$\chi^2_{thé} = \chi^2_{5-0-1} = \chi^2_4(0,05)$

$\chi^2_{thé} = 9,49$

$\chi^2_{cal} < \chi^2_{thé} \Rightarrow H_0$ acceptée au risque α